lassoRegCOVID.R

karanYsingh

2021-04-10

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.0.3
```

```
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.0.4
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-1
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
## Warning: package 'tibble' was built under R version 4.0.4
```

```
## Warning: package 'tidyr' was built under R version 4.0.4
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
## Warning: package 'purrr' was built under R version 4.0.3
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
## Warning: package 'stringr' was built under R version 4.0.3
```

```
## Warning: package 'forcats' was built under R version 4.0.3
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.3
```

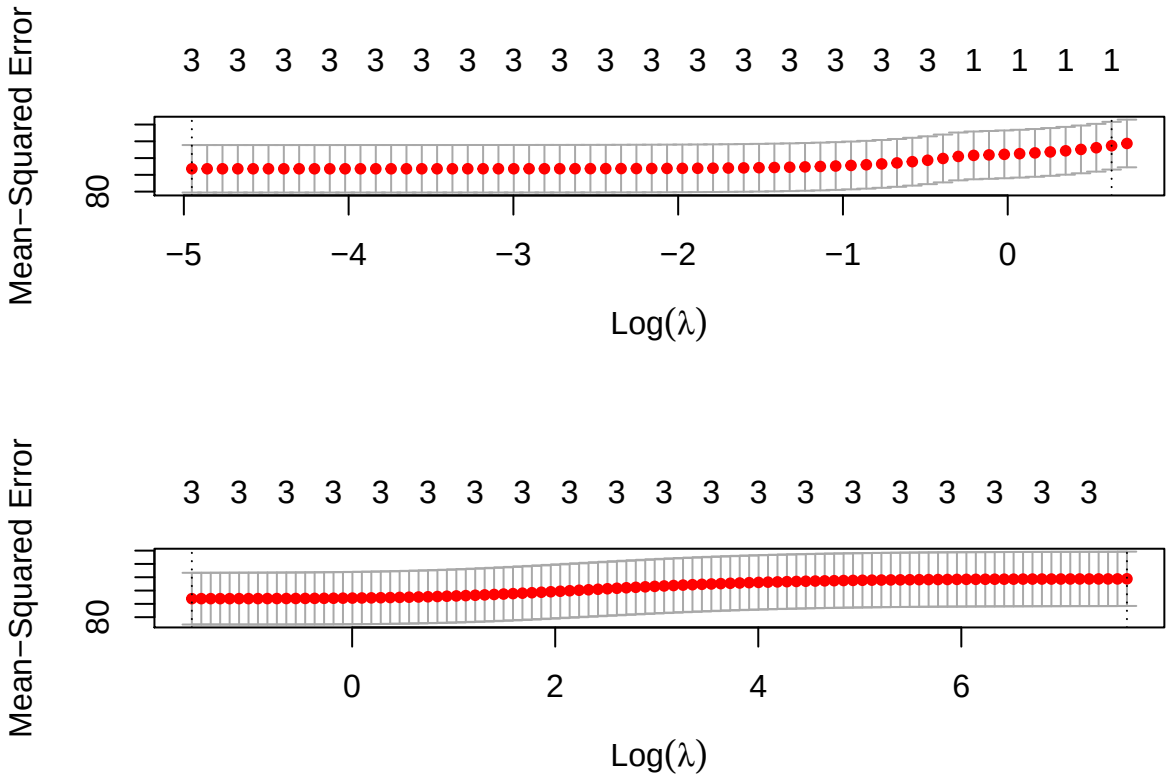
```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
data = read_xlsx("C:\\Users\\91828\\Documents\\Rlab\\Covid\\Rgeression-Model-Dataset.xlsx")
# view(data)


# Load the data and remove NAs
data <- na.omit(data)

# Split the data into training and test set
set.seed(123)
training.samples <- data$positivity_rate %>%
  createDataPartition(p = 0.8, list = FALSE)
train.data  <- data[training.samples, ]
test.data <- data[-training.samples, ]

# levels of each var[Taking only predictors with level>=2]
# map(map(train.data,as.factor),levels)
train.data <- train.data[map(map(map(train.data,as.factor),levels),length)>1]
test.data <- test.data[map(map(map(test.data,as.factor),levels),length)>1]



# Dumy code categorical predictor variables
x <- model.matrix(positivity_rate ~testing_rate+population+new_cases , train.data)


# Convert the outcome (class) to a numerical variable
y <- train.data$positivity_rate

# Find the best lambda using cross-validation
set.seed(123)
```

```
cv.lasso <- cv.glmnet(x, y, alpha = 1)
cv.ridge <- cv.glmnet(x, y, alpha = 0)
par(mfrow=c(2,1))
plot(cv.lasso)
plot(cv.ridge)
```




```
#lambda.1se ->lies within 1 standard variation of lambda.min, it gives good accuracy with minimum predi

# Fit the final model on the training data
model.lasso <- glmnet(x, y, alpha = 1,
                      lambda = cv.lasso$lambda.1se)
model.ridge <- glmnet(x,y,alpha=0,lambda = cv.ridge$lambda.1se)

# Display regression coefficients
coef(model.lasso)
```

```
## 5 x 1 sparse Matrix of class "dgCMatrix"
##                          s0
## (Intercept)  7.430311e+00
## (Intercept)  .
## testing_rate .
## population   .
## new_cases    8.891154e-06
```

```r
coef(model.ridge)
```

```
## 5 x 1 sparse Matrix of class "dgCMatrix"
##                          s0
## (Intercept)   7.494487e+00
## (Intercept)   .
## testing_rate -1.007731e-06
## population    6.282534e-13
## new_cases     4.693890e-07
```

```r
#FEVER, TIREDNESS DRYCOUGH SORETHROAT NONESYMPTON are most significant according to lasso.
#Whereas RIDGE includes all the parameters
# Make predictions on the test data LASSO
x.test <- model.matrix(positivity_rate~testing_rate+population+new_cases,test.data)[,-1]
```